





**Active gyrotactic stability of microswimmers using hydromechanical signals**Jingran Qiu <sup>1</sup>, Navid Mousavi <sup>2</sup>, Lihao Zhao <sup>1,\*</sup> and Kristian Gustavsson <sup>2,†</sup><sup>1</sup>AML, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China<sup>2</sup>Department of Physics, University of Gothenburg, SE-41296 Gothenburg, Sweden

(Received 25 May 2021; accepted 21 December 2021; published 27 January 2022)

Many plankton species undergo daily vertical migration to large depths in the turbulent ocean. To do this efficiently, the plankton can use a gyrotactic mechanism, aligning them with gravity to swim downwards or against gravity to swim upwards. Many species show passive mechanisms for gyrotactic stability. For example, bottom-heavy plankton tend to align upwards. This is efficient for upward migration in quiescent flows, but it is often sensitive to turbulence which upsets the alignment. Here we suggest a simple, robust active mechanism for gyrotactic stability, which is only lightly affected by turbulence and allows alignment both along and against gravity. We use a model for a plankton that swims with a constant speed and can actively steer in response to hydrodynamic signals encountered in simulations of a turbulent flow. Using reinforcement learning, we identify the optimal steering strategy. By using its setae to sense its settling velocity transversal to its swimming direction, the swimmer can deduce information about the direction of gravity, allowing it to actively align upwards. The mechanism leads to a rate of upward migration in a turbulent flow that is of the same order as in quiescent flows, unless the turbulence is very vigorous. In contrast, passive swimmers with typical parameters of copepods show much smaller upward velocity in turbulence. Settling may even cause them to migrate downwards in vigorous turbulence.

DOI: [10.1103/PhysRevFluids.7.014311](https://doi.org/10.1103/PhysRevFluids.7.014311)**I. INTRODUCTION**

Efficient strategies for vertical swimming under turbulence are important for many marine plankton species. Many plankton species undergo daily vertical migration of up to tens of meters to greater depths and back [1–3] to allow for efficient nutrition uptake, to avoid predators, and to adjust to tidal flows [3–6]. Many species adopt strategies leading to correlated and ballistic vertical migration along or against gravity [7,8]. Their horizontal dynamics, by contrast, is usually simply diffusive. There are several simple passive gyrotactic mechanisms to achieve rapid vertical migration. Some organisms are bottom heavy or display shape asymmetries. In both cases, the resulting torques anti-align the organism with the direction of gravity. Many plankton species also use active gyrotactic strategies. Some phytoplankton can adjust their shape to either align with or against gravity [9]. Species of protists and flagellates sense gravity using mechanosensitive ion channels to allow them to adjust their swimming direction with regards to gravity [10–12]. To

---

\*zhaolihao@mail.tsinghua.edu.cn

†kristian.gustafsson@physics.gu.se

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by [Bibsam](https://www.bibsam.com/).

our knowledge, there is no evidence that planktonic copepods sense gravity using ion channels. It has been suggested that planktonic copepods instead can use setae distributed on their body and antennae to obtain information about their orientation relative to gravity [13], but this hypothesis has not been verified or discussed in detail. The setae allow copepods to sense slip velocities as small as  $20 \mu\text{m/s}$  between themselves and the ambient fluid [14]. This is one order of magnitude smaller than their typical settling velocity in quiescent flow [15]. The setae also allow for measurement of additional hydromechanical signals such as the local strain rate and slip vorticity. The strain rate is known to be important for predation and predator avoidance [16], but the detailed response of the copepod to different hydromechanical signals remains unclear. Is it possible for copepods to use the information perceived by hydromechanical signals to achieve active gyrotaxis and, if so, what is the best swimming strategy for this? The answers to these questions may provide a new perspective to understand the vertical migration taking place in different situations [4–6]. The answers also have bearing on future applications of fabricated microswimmers.

It is not understood how microswimmers such as copepods navigate in the most efficient way in the turbulent ocean. Due to their small size and the complexity of the flow, it is difficult to perform experimental investigations on the relationship between hydromechanical signals and navigation strategies such as vertical migration. To nevertheless try to understand and explain the dynamics of plankton, a simplified model has been formulated, describing plankton as pointlike, bottom-heavy, spheroidal particles that swim with a constant speed relative to the local fluid in their instantaneous direction [17–22]. Analyzing this model using direct numerical simulations (DNS) of turbulence and analytical approaches allows one to discover the mechanisms that determine orientational statistics, clustering, preferential sampling, and vertical migration of plankton. The gyrotactic torque resulting from the bottom heaviness of the swimmer aligns it against gravity and give efficient upward vertical migration in the absence of turbulence. However, in the presence of turbulent velocity gradients, the model predicts that the upward vertical-migration velocity of copepods with parameters approximated from nature is only a fraction of the swimming speed, and upward migration may even fail due to gravitational settling [23–25]. In addition, the model does not explain how some organisms swim downwards [2,26]. Settling is too slow for copepods to migrate several meters to the depths at dawn; they must therefore use an active mechanism to descend more efficiently.

Due to the lack of experimental input, it is hard to formulate improved models where the swimmers can actively adjust their swimming behavior in response to external stimuli. One approach is to use reinforcement learning to find good strategies, allowing one to construct new models that can be compared to experiments. Proof of concept studies have shown that reinforcement learning provides strategies for efficient vertical migration [23,27–29] and more general navigation tasks [29–33] for swimmers with different shapes and motilities in both two- and three-dimensional flows. However, in these studies, the swimmers had access to global information such as their absolute position or orientation. Usually, the swimmers sense only local information, in their own frame of reference. As a consequence, vertical migration requires that external or hydrodynamic forces break the symmetry of the problem. For example, it was shown in Ref. [25] that the symmetry breaking set by gravity allows the swimmer to infer its vertical orientation by sensing only local hydromechanical signals. Using reinforcement learning, efficient steering protocols for upward navigation, based primarily on the local strain rate, were found in frozen (time-independent) two-dimensional flows. The results illustrate that the swimmers can learn rapid vertical migration by mimicking slender swimmers that tend to preferentially sample upwelling regions in the flow [20–22,34].

The mechanism mentioned above is specific to two-dimensional flows. In this paper, we therefore investigate vertical-migration strategies in three-dimensional flows. We use reinforcement learning to find candidate strategies. The best strategies found are similar to those found in the two-dimensional frozen flow [25], but it turns out that the slip velocity is the dominant signal, in contrast to the strain rate for the frozen two-dimensional flows. The main result is a simple yet powerful strategy for vertical migration based on a single component of the slip velocity, as summarized in Fig. 1. Let  $\mathbf{n}$ ,  $\mathbf{p}$ , and  $\mathbf{q}$  be an orthonormal coordinate system describing the orientation of a

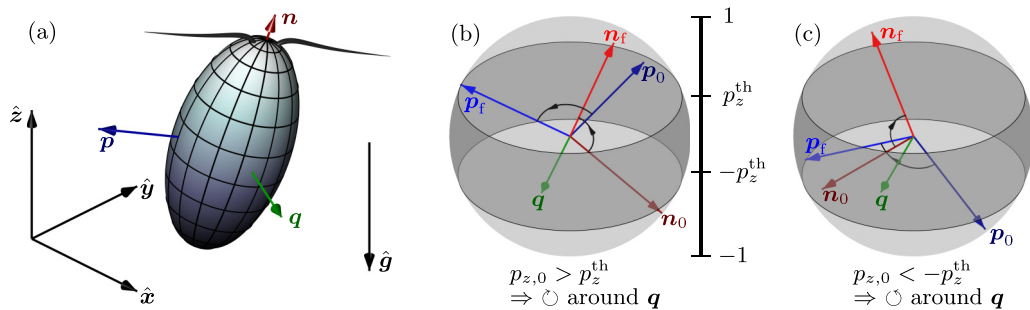


FIG. 1. (a) Orientation axes  $\mathbf{n}$ ,  $\mathbf{p}$ , and  $\mathbf{q}$  for a spheroidal swimmer in a fixed Cartesian frame of reference  $(\hat{x}, \hat{y}, \hat{z})$  with direction of gravity  $\hat{\mathbf{g}} = -\hat{z}$ . (b),(c) Unit spheres illustrating the mechanism in Eq. (11) for active gyrotactic stability. Starting with orientation  $\mathbf{n}_0$ ,  $\mathbf{p}_0$ , and  $\mathbf{q}$ , the swimmer actively rotates around its  $\mathbf{q}$  axis until  $|p_z| \leq p_z^{\text{th}}$  (shaded region) with positive rotation if (b)  $p_{z,0} > p_z^{\text{th}}$  and negative rotation if (c)  $p_{z,0} < -p_z^{\text{th}}$ , leading to the final orientation  $\mathbf{n}_f$ ,  $\mathbf{p}_f$ , and  $\mathbf{q}$  with  $n_{z,f} \geq n_{z,0}$  in general.

swimmer with  $\mathbf{n}$  in its swimming direction and  $\mathbf{p}$  in the direction perpendicular to  $\mathbf{n}$  where the swimmer has most sensitive perception of flow disturbances. It could, for instance, be the direction of the antennae of a copepod; see Fig. 1(a). The symmetry breaking due to gravitational settling allows the swimmer to measure the  $z$  component of  $\mathbf{p}$  (assuming that gravity points in the negative  $z$  direction) up to some resolution  $p_z^{\text{th}}$ . The strategy is to steer around its  $\mathbf{q}$  axis with a positive angular velocity if  $p_z > p_z^{\text{th}}$  and with a negative angular velocity when  $p_z < -p_z^{\text{th}}$ . Both cases lead to the swimmer monotonously rotating towards larger  $n_z$ . Figures 1(b) and 1(c) illustrate this strategy for the case where the swimmer is unaffected by both fluid gradients and passive gyrotactic torque due to inhomogeneous mass distribution. Independent of the initial orientation, the swimmer can monotonously rotate towards larger  $n_z$ . The strategy allows for both efficient migration upwards or downwards, not relying much on the physical characteristics of the swimmer in terms of shape, mass distribution, and threshold value of the sensing.

In Sec. II, we introduce the model for the microswimmers and our setup for reinforcement learning. In Sec. III, we show results for the mechanism described above and compare it to swimmers using passive gyrotaxis and refined active gyrotactic strategies relying on more flow signals. Conclusions and discussion of the results are presented in Sec. IV.

## II. MODEL

We use a model similar to that in Ref. [25], but here we consider a three-dimensional flow. We model the swimmer as a point particle with the shape of an elongated spheroid with aspect ratio  $\lambda = a_{\parallel}/a_{\perp}$ , where  $a_{\parallel}$  is the axis length along the symmetry direction and  $a_{\perp}$  is the transversal radius. For typical microswimmers in the ocean, inertia of both the swimmer and the fluid can be neglected [25], leading to the following dynamics for the position  $\mathbf{x}$ , the symmetry direction  $\mathbf{n}$ , the direction  $\mathbf{p}$  of the antennae, perpendicular to  $\mathbf{n}$ , and the direction  $\mathbf{q}$  perpendicular to the  $\mathbf{n}$ - $\mathbf{p}$  plane:

$$\dot{\mathbf{x}} = \mathbf{v}, \quad \dot{\mathbf{n}} = \boldsymbol{\omega} \times \mathbf{n}, \quad \dot{\mathbf{p}} = \boldsymbol{\omega} \times \mathbf{p}, \quad \mathbf{q} = \mathbf{n} \times \mathbf{p}. \quad (1a)$$

Here,  $\mathbf{v}$  and  $\boldsymbol{\omega}$  are translational and angular velocity, given by [17,25]

$$\mathbf{v} = \mathbf{u}(\mathbf{x}, t) + \mathbf{v}^{(g)} + v^{(s)}\mathbf{n}, \quad (1b)$$

$$\boldsymbol{\omega} = \boldsymbol{\Omega}(\mathbf{x}, t) + \Lambda \mathbf{n} \times [\mathbb{S}(\mathbf{x}, t)\mathbf{n}] + \frac{1}{2B}\mathbf{n} \times \hat{\mathbf{z}} + \boldsymbol{\omega}^{(s)}. \quad (1c)$$

TABLE I. Typical values of model parameters estimated from juvenile copepods in the ocean and experiments, taken from Ref. [25]. The length is obtained for small copepods [15,44] and the aspect ratio is estimated from their length and width [45]. The mass-density ratio is obtained using the mass density of copepods,  $\rho^{(p)}$ , from Ref. [46] and a sea-water density of  $\rho^{(f)} = 1.025 \text{ g/cm}^3$  obtained for 3.5% salinity at  $20^\circ\text{C}$  [47]. The settling velocity is obtained using the Stokes settling velocity for spheroids [36] with our used values for  $a_{\parallel}$ ,  $\lambda$ , and  $\rho^{(p)}/\rho^{(f)}$ . The swimming velocity is taken from the experiments in Ref. [15]. The gyrotactic reorientation time of copepods is largely unknown. The value  $B \sim 10 \text{ s}$  is taken from the critical vorticity in experiments on juvenile copepods [48]. In our simulations, we use a slightly shorter reorientation time,  $B = 5 \text{ s}$ . The maximal swimming angular velocity was estimated from the experiments in Ref. [49]. In our model, we use a smaller value that represents the slow steering motion described in Ref. [50]. The minimal sensing threshold  $\Delta u^{\text{th}} \sim 20 \text{ }\mu\text{m/s}$  is obtained from the measurements in Ref. [14]. Simulation parameters are taken as the used values.

		Range	Used value	Unit
Swimmer length	$2a_{\parallel}$	0.1–0.5	0.2	mm
Aspect ratio	$\lambda$	2.0–2.5	2.0	
Mass-density ratio	$\rho^{(p)}/\rho^{(f)}$	1.005–1.019	1.017	
Settling velocity	$v_{\parallel}^{(g)}$	0.1–0.8	0.15	mm/s
	$v_{\perp}^{(g)}$		0.13	mm/s
Swimming velocity	$v^{(s)}$	0.33–3.76	1.32	mm/s
Timescale of passive gyrotaxis	$B$	$\sim 10$	5.0	s
Swimming angular velocity	$\omega^{(s)}$	$< 20$	1.14	rad/s
Sensing threshold	$\Delta u^{\text{th}}$	$\geq 20$	50	$\mu\text{m/s}$

The flow velocity  $\mathbf{u}$ , half the flow vorticity  $\mathbf{\Omega} = \nabla \times \mathbf{u}/2$ , and the strain rate matrix  $\mathbb{S} = (\nabla \mathbf{u} + [\nabla \mathbf{u}]^T)/2$  are evaluated at the instantaneous position of the swimmer. While  $\mathbf{u}$  advects the swimmer's center of mass, the first two terms in  $\boldsymbol{\omega}$  rotate it according to Jeffery's angular velocity for a spheroid with shape parameter  $\Lambda = (\lambda^2 - 1)/(\lambda^2 + 1)$  [35].

Assuming that the direction of gravity,  $\hat{\mathbf{g}}$ , points in the negative  $\hat{\mathbf{z}}$  direction, the swimmer settles with an orientation-dependent velocity [36],

$$\mathbf{v}^{(g)} = (v_{\perp}^{(g)} - v_{\parallel}^{(g)})n_z \mathbf{n} - v_{\perp}^{(g)} \hat{\mathbf{z}}. \quad (2)$$

Here,  $v_{\parallel}^{(g)}$  and  $v_{\perp}^{(g)}$  are the velocities of a spheroidal particle settling in a quiescent flow with the symmetry axis parallel and perpendicular to gravity. Assuming a bottom-heavy mass distribution, gravity also gives rise to a passive gyrotactic torque, turning the swimmer away from  $\hat{\mathbf{g}}$  at the timescale  $B$  [37], leading to the third term in  $\boldsymbol{\omega}$ .

Finally, the last term in  $\mathbf{v}$  corresponds to swimming with constant speed  $v^{(s)}$  in the instantaneous direction  $\mathbf{n}$ , and  $\boldsymbol{\omega}^{(s)}$  in  $\boldsymbol{\omega}$  is an active angular velocity due to steering in response to measurements of the local environment. We assume that  $\omega_p^{(s)} = \boldsymbol{\omega}^{(s)} \cdot \mathbf{p}$  and  $\omega_q^{(s)} = \boldsymbol{\omega}^{(s)} \cdot \mathbf{q}$  take either of the values  $\{-\omega^{(s)}, 0, \omega^{(s)}\}$ , whereas  $\omega_n^{(s)} = \boldsymbol{\omega}^{(s)} \cdot \mathbf{n}$  remains zero. Typical dimensional parameter values of copepods are shown in Table I.

## A. Flow

We use simulations of the dynamics (1) driven by a velocity field  $\mathbf{u}$  obtained either by DNS or a statistical model for the flow velocity. To make learning fast enough during training, we use frozen flow obtained by taking snapshots of the time-dependent flows in both the DNS and the statistical model. Later, when we evaluate the found strategies, we also consider the dynamics in time-dependent flows. This approach of training in frozen flows only takes into account the spatial structure of the flow. We do not address the question of whether there exist even better strategies that exploit spatiotemporal structures in turbulence.

In our DNS, we use the method of Ref. [24] to simulate incompressible homogeneous isotropic turbulence by numerical solution of Navier-Stokes equations,

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{\nabla p^{(f)}}{\rho^{(f)}} + \nu \nabla^2 \mathbf{u} + \mathbf{f}, \quad \nabla \cdot \mathbf{u} = 0, \quad (3)$$

on a periodic domain. Here,  $p^{(f)}$ ,  $\rho^{(f)}$ , and  $\nu$  denote the pressure, density, and kinematic viscosity of the fluid, respectively. The turbulence is driven by an external force  $\mathbf{f}$  with low wave number [38]. Equations (3) are solved by a pseudospectral method, and the 3/2 rule is adopted to reduce the aliasing error on the nonlinear term. Time integration of both fluid and swimmer dynamics is performed using the explicit second-order Adam-Bashforth scheme [39]. Our simulations are carried out using  $96^3$  grid points with Taylor scale Reynolds number  $\text{Re} \approx 60$ . The smallest resolved scale is about 1.78 times smaller than the Kolmogorov scale, which lies within the accepted range to reach statistically reliable results [40].

In the statistical model, we use an incompressible Gaussian random velocity field of the form

$$\mathbf{u}(\mathbf{x}, t) = \frac{1}{\sqrt{6}} \nabla \times \mathbf{A}(\mathbf{x}, t), \quad (4)$$

where  $\mathbf{A}$  is a three-dimensional vector potential with components being independent Gaussian random functions with zero mean and correlation function [41]

$$\langle A_i(\mathbf{x}, t) A_j(\mathbf{x}', t') \rangle = \delta_{ij} (\ell u^{\text{rms}})^2 \exp \left[ -\frac{|\mathbf{x}' - \mathbf{x}|^2}{2\ell^2} - \frac{|t - t'|}{\tau} \right]. \quad (5)$$

Here,  $\langle \cdot \rangle$  denotes a steady-state ensemble average,  $\ell$  and  $\tau$  are the characteristic length scale and timescale, and  $u^{\text{rms}} = \langle \mathbf{u}^2 \rangle^{1/2}$ . The statistical model shows the best agreement with simulations in turbulent flows if large but finite values of  $\text{Ku} = u^{\text{rms}} \tau / \ell$  are considered. In this limit, the flow evaluated at the positions of Lagrangian tracer particles decorrelates on the timescale  $\ell / u^{\text{rms}}$  due to displacement, rather than the timescale  $\tau$  due to temporal variations of the flow. The relevant timescale in the statistical model in the limit of large  $\text{Ku}$  is thus proportional to the Kolmogorov time  $\tau^{(\eta)} = (\text{tr}(2\mathbb{S}^T \mathbb{S}))^{-1/2}$ , where  $\mathbb{S}$  is the strain rate matrix. In the statistical model, the Kolmogorov time evaluates to  $\tau^{(\eta)} = \ell / (\sqrt{5} u^{\text{rms}})$  [41].

The root-mean-square velocity of turbulent fluctuations in the ocean,  $u^{\text{rms}}$ , ranges from 0.1 to 100 mm/s [42]. The energy dissipation rate ranges from  $10^{-4}$  mm<sup>2</sup>/s<sup>3</sup> in the deep sea, up to 100 mm<sup>2</sup>/s<sup>3</sup> in the upper ocean mixing layer [42,43]. For a kinematic viscosity of  $\nu = 1$  mm<sup>2</sup>/s, the Kolmogorov time  $\tau^{(\eta)}$  takes values between 0.1 and 100 s.

## B. Sensing and actions

Small microorganisms equipped with setae can use them to measure velocity differences between their body and the surrounding fluid [13,14]. Experiments show that copepods respond in different ways depending on the magnitudes of strain rate  $\mathbb{S}$ , angular velocity differences  $\Delta \boldsymbol{\Omega}$ , and velocity differences  $\Delta \mathbf{u}$  of the flow [51], indicating that they can distinguish between these three signals while not swimming. This ability could be a consequence of that different flow structures give rise to different bending patterns of the copepod's setae [51,52]. The disturbance to the surrounding flow by swimming complicates measurements [52]. However, copepods are able to distinguish external hydrodynamic signals from their own generated flow when feeding [53]. It is therefore plausible that copepods are able to distinguish external hydromechanical signals from their own flow during steady swimming by recognizing spatial and temporal flow structures [54] using an array of densely distributed setae along their body and antennae [55]. Copepods often show a directional bias in their ability to sense hydromechanical signals [56]. We assume a microorganism that is better at detecting signals in the symmetry direction  $\mathbf{n}$  and the directions of its antennae  $\pm \mathbf{p}$  than in the direction  $\mathbf{q}$ . Primarily, we therefore consider navigation using the signals  $\mathbb{S}$ ,  $\Delta \boldsymbol{\Omega}$ , and  $\Delta \mathbf{u}$ , projected on the  $\mathbf{n}$  and  $\mathbf{p}$  directions in the local frame of the swimmer. For completeness, we also discuss strategies

including the  $\mathbf{q}$  direction because, in principle, a generic swimmer, which is able to use its setae to measure the flow at enough independent positions and orientations, has enough information to distinguish signals along both directions  $\mathbf{p}$  and  $\mathbf{q}$ .

To reduce the number of signals, we use Eqs. (1) to express velocity and angular velocity differences in the local coordinate system  $\mathbf{n}$ ,  $\mathbf{p}$ , and  $\mathbf{q}$ ,

$$\Delta u_n = u_n - v_n = v_{\parallel}^{(g)} n_z - v^{(s)}, \quad (6a)$$

$$\Delta u_p = u_p - v_p = v_{\perp}^{(g)} p_z, \quad (6b)$$

$$\Delta u_q = u_q - v_q = v_{\perp}^{(g)} q_z, \quad (6c)$$

$$\Delta \Omega_n = \Omega_n - \omega_n = -\omega_n^{(s)}, \quad (6d)$$

$$\Delta \Omega_p = \Omega_p - \omega_p = \Lambda S_{nq} + \frac{1}{2B} q_z - \omega_p^{(s)}, \quad (6e)$$

$$\Delta \Omega_q = \Omega_q - \omega_q = -\Lambda S_{np} - \frac{1}{2B} p_z - \omega_q^{(s)}. \quad (6f)$$

Here, subscripts  $n$ ,  $p$ , or  $q$  denote the scalar product with the corresponding unit vector  $\mathbf{n}$ ,  $\mathbf{p}$ , or  $\mathbf{q}$ . First, according to Eq. (6a), the magnitude  $|\Delta u_n| \sim v^{(s)}$  is much higher than our used sensing threshold in Table I. To avoid introducing additional arbitrary threshold values, we skip this signal. Second, Eqs. (6b) and (6c) show that the swimmer is able to directly measure the  $z$  component of the  $\mathbf{p}$  and  $\mathbf{q}$  directions and we adopt these components as signals. Third, Eq. (6d) shows that measurement of  $\Delta \Omega_n$  does not provide any information. Moreover, Eqs. (6e) and (6f) show that  $\Delta \Omega_p$  is given by the signals  $S_{nq}$  and  $q_z$ , and  $\Delta \Omega_q$  is given by  $S_{np}$  and  $p_z$ . Therefore,  $\Delta \Omega$  does not contribute with any independent information and we neglect it as a signal. Fourth, we only consider the strain components that directly affect the dynamics in Eqs. (6),  $S_{np}$  and  $S_{nq}$ . Other strain components may give relevant information due to flow correlations, but this is most likely secondary to the direct contributions of  $S_{np}$  and  $S_{nq}$ . Finally, in our model, we have omitted higher-order derivatives and time dependence of the hydromechanical signals.

The aim is to use reinforcement learning to search for optimal or approximately optimal strategies for vertical migration by suitable steering (actions) based on hydromechanical signals (states). In our model, we use different sets of states obtained from combinations of discretized signals  $p_z$ ,  $q_z$ ,  $S_{np}$ , and  $S_{nq}$ . Each signal  $\sigma$  is discretized into three states separated by a threshold level  $\sigma^{\text{th}}$ ,

$$\text{states for signal } \sigma = \begin{cases} \sigma^{(+)} & \text{if } \sigma > \sigma^{\text{th}} \\ \sigma^{(0)} & \text{if } |\sigma| < \sigma^{\text{th}} \\ \sigma^{(-)} & \text{if } \sigma < -\sigma^{\text{th}}. \end{cases} \quad (7)$$

For the velocity differences  $\Delta u_p$  and  $\Delta u_q$ , we use a threshold  $\Delta u^{\text{th}} = 50 \mu\text{m/s}$ , giving the threshold  $p_z^{\text{th}} = \Delta u^{\text{th}}/v_{\perp}^{(g)} \approx 0.37$  for the  $p_z$  and  $q_z$  components with  $v_{\perp}^{(g)}$  from Table I. Estimating the corresponding gradients along the length of the swimmer,  $\Delta u^{\text{th}}/a_{\parallel}$  [16], we obtain the threshold we use for angular velocity differences and strain rates,  $S^{\text{th}} = 0.5 \text{ s}^{-1}$ . We have chosen the threshold  $\Delta u^{\text{th}}$  to lie between the smallest velocity difference that copepods can physically sense,  $20 \mu\text{m/s}$  [14], and the settling velocity  $\sim 150 \mu\text{m/s}$  in order to allow the copepod to sense its settling. In experiments, it is observed that different species of copepods make vigorous escape jumps in response to steady-flow strain rates of the order of  $0.2\text{--}20 \text{ s}^{-1}$  [16,44,51,57]. Our threshold  $S^{\text{th}}$  is of this order of magnitude, indicating that it can be measured by copepods in nature. We have confirmed that our results described in the next section are not sensitive to the exact values of the thresholds, as long as  $\Delta u^{\text{th}}$  is smaller than half the settling velocity.

We assume that the active angular velocity contribution,  $\omega^{(s)}$  in Eq. (1c), allows the swimmer to steer by rotating around the  $\mathbf{p}$  and  $\mathbf{q}$  axes with angular velocities  $\omega_p^{(s)}$  and  $\omega_q^{(s)}$ , respectively. We first consider a planar model, where the swimmer can measure signals in the  $\mathbf{n}\text{-}\mathbf{p}$  plane:  $p_z$

and  $S_{np}$ . In this case, we assume no steering around the  $\mathbf{p}$  axis,  $\omega_p^{(s)} = 0$ , and the angular velocity  $\omega_q^{(s)}$  around the  $\mathbf{q}$  axis takes three values,  $\omega_q^{(s)} = \{-\omega^{(s)}, 0, +\omega^{(s)}\}$ . This planar model is similar to the two-dimensional model considered in Ref. [25], but in that case,  $\mathbf{q}$  was fixed to the direction perpendicular to the flow plane, while  $\mathbf{q}$  is free to rotate in the three-dimensional flow considered here. We also consider a full three-dimensional model, where the swimmer in addition to the signals of the planar model has access to the signals  $S_{nq}$  and  $q_z$ . In addition to the three values of angular velocities around the  $\mathbf{q}$  axis, the swimmer in the three-dimensional (3D) model can also steer around the  $\mathbf{p}$  axis with three levels  $\omega_p^{(s)} = \{-\omega^{(s)}, 0, +\omega^{(s)}\}$ , giving, in total, nine different actions.

Copepods can acquire angular velocities up to 20 rad/s before a jump [49]. When cruising, they steer with smaller angular velocities [50]. In our reinforcement learning, we use  $\omega^{(s)} = 1.14$  rad/s. We choose this value because it is of the order of the root-mean-square vorticity in our simulations during training, and it is much smaller than the hypothetical maximal angular velocity,  $v^{(s)}/a_{\parallel} \approx 13$  rad/s, that would have been obtained if the swimmer were able to convert the full swimming propulsion into angular rotation. Our value  $\omega^{(s)} = 1.14$  rad/s gives a large length-specific turning radius,  $v^{(s)}/(\omega^{(s)}a_{\parallel}) \approx 12$ , consistent with slow steering while cruising. We have confirmed that our results described in the next section are not sensitive to the exact value of  $\omega^{(s)}$ , as long as it is not too small.

### C. Reinforcement learning for vertical migration

We adopt a one-step  $Q$ -learning algorithm [58–60] to train the swimmer to find efficient strategies for vertical migration. The training is divided into a number of episodes. In each episode, one precalculated frozen flow snapshot is randomly chosen. The swimmer starts with a random initial position and orientation and follows the dynamics (1), first for an equilibration time  $T^{\text{eq}}$  and then for a predefined physical time of training,  $T^{\text{train}}$ . The initial equilibration is introduced to search for strategies that are efficient in the statistical steady state, not relying on the arbitrary uniform initial condition. In the DNS at regularly distributed time steps  $iT^{\text{meas}}$  with  $i = 0, 1, \dots, \lfloor (T^{\text{eq}} + T^{\text{train}})/T^{\text{meas}} \rfloor$ , the swimmer measures its current discrete state  $s_i$ , obtained by measurement of the signals described above, and is given a reward  $r_i$ . In the statistical model, a reward is instead given at any time step where the state changes. Because the goal is vertical migration upwards, we use the velocity in the  $z$  direction since the last state measurement as a reward in the DNS,  $r_i = (z_{i+1} - z_i)/T^{\text{meas}}$ , and displacement,  $r_i = z_{i+1} - z_i$ , in the statistical model. After the reward is given, the swimmer sets  $\omega^{(s)}$  to action  $a_i$ . During the equilibration,  $T < T^{\text{eq}}$ , the action is chosen according to a greedy policy,  $a_i = \arg \max_a Q(s_i, a)$ , and the  $Q$ -table is kept constant. After the equilibration, the action is chosen according to an  $\varepsilon$ -greedy policy,

$$a_i = \begin{cases} \text{random action} & \text{with probability } \varepsilon \\ \arg \max_a Q(s_i, a) & \text{otherwise} \end{cases}, \quad \text{where } \varepsilon = \varepsilon_0 \max\left(0, 1 - \frac{E}{E^{(\varepsilon)}}\right). \quad (8)$$

Here,  $\varepsilon$  is a small exploration rate that starts at  $\varepsilon_0$  and decays linearly with the training episode number  $E$  until it reaches zero at episode  $E^{(\varepsilon)}$ . It allows the swimmer to explore different actions, preventing the learning to get stuck at the local optima. Normally, nonexploratory actions are chosen as the action giving the maximal value of the  $Q$ -table,  $Q(s, a)$ , for the current state. Starting with a  $Q$ -table of uniform values, it is updated each time the agent is given a reward using the rule

$$Q(s_i, a_i) \leftarrow Q(s_i, a_i) + \alpha [r_i + \gamma \max_a Q(s_{i+1}, a) - Q(s_i, a_i)] \quad \text{where } \alpha = \alpha_0 \frac{E^{(\alpha)}}{E^{(\alpha)} + E}. \quad (9)$$

Here,  $\alpha$  is a learning rate, starting at  $\alpha_0$  and decaying on the episode scale  $E^{(\alpha)}$ . The discount factor  $\gamma$ , with  $0 \leq \gamma < 1$ , sets an optimization timescale over  $(1 - \gamma)^{-1}$  state changes. It is introduced to prevent divergence of the values in the  $Q$ -table in the long run. We adopt a farsighted optimization by setting  $\gamma$  close to 1. At the end of each episode, the  $Q$ -table is kept for the next episode, continuing to be updated. In the ideal case of infinitely many exploratory moves and if the system is Markovian, the entries of the  $Q$ -table approach the optimal values for the expectation values of the future

TABLE II. Average vertical velocity  $\langle v_z \rangle$  normalized by the value in quiescent flows  $\langle v_z \rangle_0$  [Eq. (10)] for different sets of discretized actions and states (see Sec. II B), with parameters from Table I. Numbers show results obtained by evaluation of the best strategy found for each case using reinforcement learning in frozen DNS or statistical model flows. The numbers in parentheses show results for the strategies in Eqs. (11), (14), and (15). Errors of the numerical values are of the order of  $\pm 0.01$ .

Model	Actions	Signals	Performance $\langle v_z \rangle / \langle v_z \rangle_0$				
			Frozen DNS		Frozen statistical model		
Naive			0.30		0.22		
Planar	$\omega_q^{(s)}$	$p_z$	0.58	(0.58)	0.73	(0.73)	[Eq. (11)]
”	”	$S_{np}$	0.31		0.22		
”	”	$p_z, S_{np}$	0.58		0.73		
Full 3D	$\omega_p^{(s)}, \omega_q^{(s)}$	$p_z$	0.58		0.73		
”	”	$q_z$	0.58		0.73		
”	”	$p_z, q_z$	0.64	(0.66)	0.94	(0.94)	[Eq. (14)]
”	”	$S_{np}$	0.32		0.22		
”	”	$S_{nq}$	0.34		0.22		
”	”	$S_{np}, S_{nq}$	0.31		0.20		
”	”	$p_z, S_{np}$	0.57		0.73		
”	”	$p_z, S_{nq}$	0.58		0.73		
”	”	$q_z, S_{np}$	0.57		0.73		
”	”	$q_z, S_{nq}$	0.58		0.73		
”	”	$p_z, q_z, S_{np}, S_{nq}$	0.64	(0.68)	0.94	(0.94)	[Eq. (15)]

discounted reward,  $Q(s_i, a) = \langle \sum_{j=0}^{\infty} \gamma^j r_{i+j} \rangle$ , for taking action  $a$  in state  $s_i$  following the policy (8). In most realistic situations, the process instead converges to an approximately optimal strategy. Information about the parameters used in our different training cases are given in Appendix A.

### III. OPTIMAL SWIMMING STRATEGIES

#### A. Results from reinforcement learning

We evaluate the performance of the strategies found in our reinforcement learning by running the same setup as in the training, but with a greedy action ( $\varepsilon = 0$ ), and we do not update the  $Q$ -table ( $\alpha = 0$ ). We average the vertical velocity over multiple episodes: each of time  $T^{\text{train}}$  sampled after the initial equilibration time  $T^{\text{eq}}$ . Both  $T^{\text{eq}}$  and  $T^{\text{train}}$  are chosen much larger than the maximal timescale of the dynamics, so that steady-state statistics is obtained. The results of the best strategies found are summarized in Table II. The vertical velocity is normalized using the average

$$\langle v_z \rangle_0 = v^{(s)} - v_{\parallel}^{(g)}, \quad (10)$$

obtained in the long-time limit if the ambient flow velocity  $\mathbf{u}$  is put to zero in Eq. (1b). Results for the naive model, obtained for a swimmer with zero active angular velocity,  $\omega^{(s)} = 0$ , are given as a reference.

In the planar model, the swimmer can steer around the  $\mathbf{q}$  axis,  $\omega^{(s)} = \omega_q^{(s)} \mathbf{q}$ , and it can sense combinations of two signals: the vertical component  $p_z$  of the  $\mathbf{p}$  axis and the strain component  $S_{np}$ . The results in Table II show that the signal  $p_z$  is more important for upward vertical migration than  $S_{np}$  in both the DNS and the statistical model. The optimal strategy when using only  $p_z$  as the signal is the same in the two flows, and it is highlighted using red frames in Figs. 2(a) and 2(b). Figure 1 shows the mechanism explaining why this strategy is successful. The optimal strategy when using only  $S_{np}$  as the signal is on the same level as the naive strategy. The optimal strategy using both



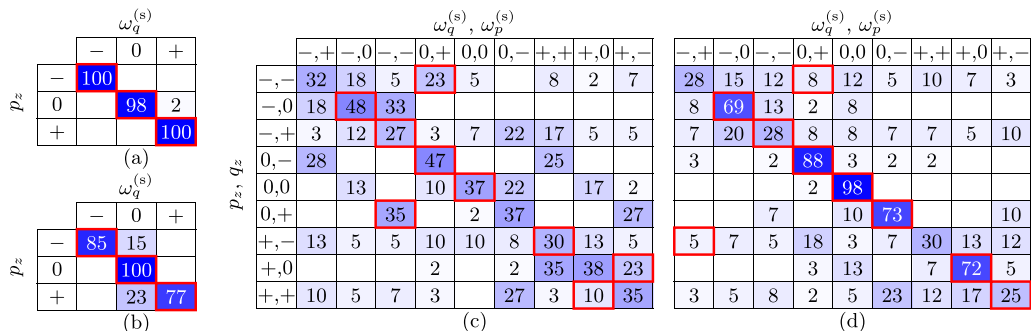


FIG. 2. Summary of strategies obtained by reinforcement learning for the cases of (a), (b) the planar model with single signal  $p_z$  and (c), (d) full 3D model with joint signal  $p_z$  and  $q_z$ . Results from (a), (c) the frozen DNS and (b), (d) the statistical model. Rows represent different states; the signs  $-$ ,  $0$ , and  $+$  represent the states  $\sigma^{(-)}$ ,  $\sigma^{(0)}$ , and  $\sigma^{(+)}$  in the discretized signal in Eq. (7). Columns represent different actions, where the signs represent the values  $-\omega^{(s)}$ ,  $0$ , and  $+\omega^{(s)}$ , respectively. The action selected in a given state for the best strategy found in our reinforcement learning is highlighted in red. In each panel, the resulting strategies from 60 completed training sessions are summarized. Each numerical value gives the nonzero percentage of strategies choosing a certain action in a given state. The background is color coded according to this percentage from white (0%) to blue (100%).

$p_z$  and  $S_{np}$  performs approximately on the same level as the simpler strategy using only  $p_z$  as the signal.

In the full 3D model, the swimmer can rotate around both the  $\mathbf{p}$  and  $\mathbf{q}$  axes, and all four relevant flow signals are considered. Using a single signal  $p_z$ , the best strategy found for both the DNS and statistical model is the same as in the planar case, i.e., rotate around the  $\mathbf{q}$  axis according to Fig. 2(a), with no rotation around the  $\mathbf{p}$  axis. Using  $q_z$  as the signal, the equivalent strategy is found by rotating around  $\mathbf{p}$  with no rotation around  $\mathbf{q}$ . For two combined signals, the best strategies found in both the DNS and statistical model are based on  $p_z$  and  $q_z$ . They are highlighted using red frames in Figs. 2(c) and 2(d).

The strategies based only on the strain signals perform on the level of the naive swimmer, with approximately half the vertical velocity compared to the strategies based on  $p_z$  and  $q_z$ . Hybrid strategies between one strain component and either of  $p_z$  or  $q_z$  perform on the same level as the individual  $p_z$  or  $q_z$  signal. Finally, in the case of four signals,  $p_z$ ,  $q_z$ ,  $S_{np}$ , and  $S_{nq}$ , the best strategies found are of the same order as the strategies based on  $p_z$  and  $q_z$  only.

Being a stochastic process, different training sessions often converge to different approximately optimal strategies. Therefore, we have run several training sessions for each case considered, and the results displayed in Table II are based on the strategy with the best performance during training. The cell colors in Fig. 2 summarize the resulting strategies from 60 training sessions for each case [Figs. 2(a)–2(d)]. For a given state, an action with a white or light-blue cell is rarely chosen, while actions with blue color are selected in most of the resulting strategies. We find that for a single signal in the planar model, most training sessions converge to the optimal strategy, highlighted in red in Figs. 2(a) and 2(b). For two signals in the full 3D model, results are more scattered. The reason is that many strategies have similar reward levels and it is therefore hard to find the global optimal strategy. In the DNS, the mean value of  $\langle v_z \rangle / \langle v_z \rangle_0$  obtained in the 60 training sessions is 0.60 and the median is 0.61, both being close to the value 0.64 for the best strategy found. The same holds for the statistical model where the mean value is 0.85 and the median is 0.94, equal to the value for the best strategy found. We conclude that although it is hard to find the global optimum, many quasioptimal strategies have approximately the same performance.

The strategies discussed above are obtained in frozen flows. The flow is changed each episode to reduce bias towards specific frozen flows. However, the flow statistics of the dynamics in a frozen

flow is slightly different from the statistics in time fluctuating flows, or for swimmers with other parameter values. In general, which strategy is optimal for a given task depends on both the model parameters and learning parameters, such as the choice of actions and states. There is no guarantee that a strategy found for one setup is also the best strategy for a different setup. For example, the optimal strategy found using a frozen flow is not necessarily the same as the optimal strategy in a time-dependent flow, and a flow in nature could have yet another optimal strategy. We therefore try to identify generally valid strategies that are robust upon changing the details. Using the trend that the best strategies in Fig. 2 tend to choose actions along the diagonals, we formulate strategies below in the planar model and full 3D model that have good performance for a large number of parameters and flows, including time-dependent flows.

### B. Planar model with a single signal

Our reinforcement learning shows that for upward vertical navigation in the planar model using a single observable and the action to steer around the  $\mathbf{q}$  axis, the most important signal is  $p_z$ , being proportional to the cosine of the angle between the transversal direction  $\mathbf{p}$  and the direction of gravity  $\hat{\mathbf{g}} = -\hat{\mathbf{z}}$ . The resulting optimal strategy, shown in Figs. 2(a) and 2(b), can be written mathematically as  $\boldsymbol{\omega}^{(s)} = \omega_p^{(s)} \mathbf{p} + \omega_q^{(s)} \mathbf{q}$ , with

$$\omega_p^{(s)} = 0 \quad \text{and} \quad \omega_q^{(s)} = \omega^{(s)} \begin{cases} 1 & \text{if } p_z > p_z^{\text{th}} \\ -1 & \text{if } p_z < -p_z^{\text{th}} \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

This strategy imposes an active gyrotactic stability on the orientational dynamics. To understand the mechanism of Eq. (11), we consider a simplified dynamics where passive gyrotaxis and flow gradients are neglected. In this limit, the angular velocity of the swimmer, given by Eq. (1c), is solely determined by the active swimming contribution,  $\boldsymbol{\omega} = \boldsymbol{\omega}^{(s)}$ , and the orientational dynamics of Eq. (1a) becomes

$$\dot{\mathbf{n}} = \omega_q^{(s)} \mathbf{p}, \quad \dot{\mathbf{p}} = -\omega_q^{(s)} \mathbf{n}, \quad \text{and} \quad \dot{\mathbf{q}} = 0. \quad (12)$$

Assuming  $\omega^{(s)} > 0$ , the signature of  $\omega_q^{(s)}$  only depends on  $p_z$  according to Eq. (11). If the initial value  $p_{z,0}$  is smaller than the sensing threshold,  $|p_{z,0}| \leq p_z^{\text{th}}$ , the orientation remains unchanged. If, instead,  $|p_{z,0}| > p_z^{\text{th}}$ , the active rotational swimming turns the swimmer around its  $\mathbf{q}$  axis until  $p_z$  reaches  $p_z^{\text{th}}$  for the case of positive  $p_{z,0}$ , or until it reaches  $-p_z^{\text{th}}$  for the case of negative  $p_{z,0}$ . The direction of the rotation following from Eq. (11) is such that the  $z$  component of  $\mathbf{n}$  in Eq. (12) is rotated towards larger values,  $\dot{n}_z = \omega^{(s)} |p_z| \Theta(|p_z| - p_z^{\text{th}}) \geq 0$ , where  $\Theta$  is the Heaviside step function. Figures 1(b) and 1(c) illustrate this mechanism for the case  $p_z^{\text{th}} = 0.37$ .

To evaluate the average alignment  $\langle n_z \rangle$  for the simplified dynamics (12), we use the normalization  $n_z^2 + p_z^2 + q_z^2 = 1$ . If  $|p_{z,0}| \leq p_z^{\text{th}}$ , there is no rotation and the final value  $n_{z,f}$  is equal to the initial value  $n_{z,0}$ . If  $|p_{z,0}| > p_z^{\text{th}}$ , then  $q_z$  remains constant, while  $p_z$  is rotated until the threshold  $\pm p_z^{\text{th}}$ .

Using the normalization, the final value of  $n_z$  must take the form  $n_{z,f} = \pm \sqrt{1 - [p_z^{\text{th}}]^2 - q_{z,0}^2}$ . The sign choice in Eq. (11) breaks symmetry, always giving the positive solution of  $n_{z,f}$ , while the negative solution would be obtained by rotating in the opposite direction of Eq. (11). Averaging the positive solution  $n_{z,f}$  over initially uniformly distributed orientations, assuming  $0 \leq p_z^{\text{th}} \leq 1$ , gives (see Appendix B for details)

$$\langle n_z \rangle = \frac{4}{3\pi} \frac{(1 - [p_z^{\text{th}}]^2)^{3/2}}{p_z^{\text{th}}} {}_3F_2 \left[ \frac{1}{2}, 1, 2; \frac{3}{2}, \frac{5}{2}; 1 - \frac{1}{[p_z^{\text{th}}]^2} \right]. \quad (13)$$

Here,  ${}_3F_2$  is the generalized hypergeometric function. This solution scales as  $\langle n_z \rangle \sim \frac{\pi}{4} - \frac{2}{\pi} p_z^{\text{th}} - \frac{\pi}{4} [p_z^{\text{th}}]^2$  for small  $p_z^{\text{th}}$ . The limiting case of very high resolution of the signal,  $p_z^{\text{th}} \rightarrow 0$ , results in

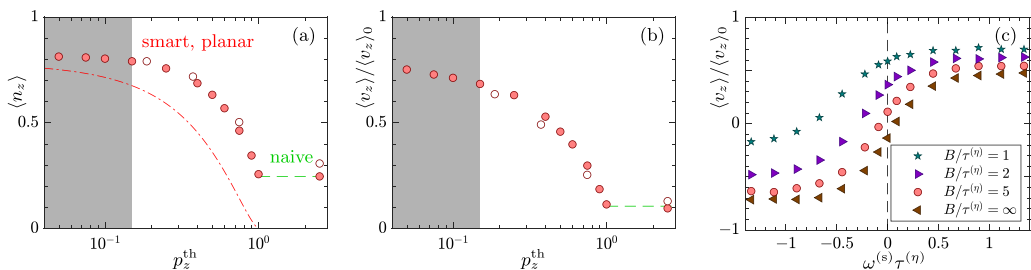


FIG. 3. Results of numerical simulations in the steady state for (a) average alignment and (b), (c) velocity against (a), (b) the threshold of the flow signal,  $p_z^{\text{th}} = \Delta u^{\text{th}}/v_{\perp}^{(g)}$ , and (c) the dimensionless angular swimming velocity  $\omega^{(s)}\tau^{(\eta)}$ . Velocities are normalized using  $\langle v_z \rangle_0$  in Eq. (10). Symbols show results from simulations of Eqs. (1) with smart planar steering [ $\omega^{(s)} = \omega_q^{(s)}\mathbf{q}$  using Eq. (11)] in the statistical model with  $\text{Ku} = 10$  (colored symbols) and DNS [hollow symbols in (a) and (b)]. Numerical results for the naive strategy ( $\omega^{(s)} = 0$ ) in the statistical model are shown as horizontal dashed green lines. The shaded region corresponds to threshold levels below the minimal resolution limit,  $\Delta u = 20 \mu\text{m/s}$ . The dash-dotted line in (a) shows the theory for the simplified dynamics, given by Eq. (13). Flow parameters are  $u^{\text{rms}} = 6.7 \text{ mm/s}$  and  $\tau^{(\eta)} = 1 \text{ s}$ . All swimmer parameters are according to the values used in Table I, except for parameter values stated in the figure.

large, albeit not perfect, alignment,  $\langle n_z \rangle \rightarrow \frac{\pi}{4} \approx 0.8$ . As  $p_z^{\text{th}}$  approaches unity, the solution scales as  $\langle n_z \rangle \sim \frac{8\sqrt{2}}{3\pi}(1 - p_z^{\text{th}})^{3/2}$ . For  $p_z^{\text{th}} > 1$ , the swimmer cannot resolve  $p_z$  and active alignment fails. We conclude that Eq. (12) is a simple mechanism for active gyrotactic stability, leading to partial alignment against gravity in quiescent flows even without inhomogeneous mass distribution.

Figure 3 shows simulation results for the full dynamics (1) using the strategy (11) with parameters from Table I. The data illustrate the sensitivity of the strategy to the choice of the dimensionless parameters  $p_z^{\text{th}}$ ,  $\omega^{(s)}\tau^{(\eta)}$ , and  $B/\tau^{(\eta)}$ . Figure 3(a) shows the dependence of the average vertical alignment  $\langle n_z \rangle$  on the signal threshold  $p_z^{\text{th}}$ . As expected, for  $p_z^{\text{th}} > 1$ , the alignment is equal to the naive case with no steering,  $\omega^{(s)} = 0$ . In this case, there is a weak alignment due to the passive gyrotactic angular velocity in Eq. (1c). When  $p_z^{\text{th}}$  is reduced below unity, there is a sharp increase in  $\langle n_z \rangle$  until  $p_z^{\text{th}} \approx 0.5$ . For smaller values of  $p_z^{\text{th}}$ ,  $\langle n_z \rangle$  is approximately constant around 0.8. We conclude that the results are not sensitive to the value of  $p_z^{\text{th}}$ , as long as it is smaller than 0.5. Our choice of  $p_z^{\text{th}} = 0.37$  lies in the upper part of this range. The dash-dotted curve in Fig. 3(a) shows the analytical evaluation for a quiescent flow, given by Eq. (13). It has the same trend as the numerical data, but predicts a somewhat smaller alignment. This is expected because the passive gyrotactic contribution in the numerical simulations adds to the alignment and the turbulent velocity gradients do not affect the alignment too much for the flow with  $u^{\text{rms}} = 6.7 \text{ mm/s}$  and  $\tau^{(\eta)} = 1 \text{ s}$  in Fig. 3. Figure 3(b) shows that the same conclusion holds for the average vertical velocity  $\langle v_z \rangle$ . Starting at the naive result for  $p_z^{\text{th}} > 1$ ,  $\langle v_z \rangle$  increases quickly until  $p_z^{\text{th}} \sim 0.5$ , where a plateau is reached with velocity around  $0.7\langle v_z \rangle_0$ , with  $\langle v_z \rangle_0$  given in Eq. (10). Hollow markers in Figs. 3(a) and 3(b) show results from our DNS. Since they qualitatively agree with the statistical model, we limit the analysis in what follows to the statistical model, which is quicker to evaluate.

Figure 3(c) shows the vertical velocity component for statistical model simulations against  $\omega^{(s)}\tau^{(\eta)}$  for different values of  $B/\tau^{(\eta)}$ . The data reach plateaus for  $|\omega^{(s)}\tau^{(\eta)}| > 0.8$ , meaning that the results are not sensitive to the choice of the dimensionless steering angular velocity if it is large enough. For  $\tau^{(\eta)} = 1 \text{ s}$ , the choice  $\omega^{(s)}\tau^{(\eta)} = 1.14$  in Table I lies well within the plateaus. When  $B/\tau^{(\eta)} \rightarrow \infty$  (brown  $\triangleleft$ ), the passive gyrotactic torque vanishes, and by solely using the active gyrotactic mechanism (11), the swimmer reaches an upward velocity of about  $0.5\langle v_z \rangle_0$ . For negative values of  $\omega^{(s)}$ , the active mechanism works in the opposite direction, giving alignment with gravity and a means to migrate downwards with a slightly larger speed due to the contribution from settling. Adding passive gyrotaxis with reorientation time  $B/\tau^{(\eta)} = 5$  from Table I gives an

additive contribution of around  $0.1\langle v_z \rangle_0$  for the full range of  $\omega^{(s)}\tau^{(\eta)}$  (red  $\circ$ ). To modify the vertical velocity significantly, the reorientation time must be of the order of  $\tau^{(\eta)}$  (green  $\star$ ) for the flow considered here. Results for passive gyrotaxis are obtained along the line  $\omega^{(s)} = 0$  in Fig. 3(c). When  $B/\tau^{(\eta)} > 2$ , the active gyrotactic stability is dominant, while for  $B/\tau^{(\eta)} < 2$ , passive gyrotaxis is efficient for upward migration and the active mechanism only gives a minor contribution.

In conclusion, our numerical simulations show that the good performance of the found strategy is not sensitive to the precise values of the parameters or the statistics of the flow used in training. For the parameters considered in Fig. 3, vertical alignment due to passive and active gyrotactic reorientation results in high vertical velocities. Both the average flow velocity  $\langle u_z \rangle$  and the settling velocity  $\mathbf{v}^{(g)}$  in Eq. (1b) are, at most, of the order of  $0.1\langle v_z \rangle_0$ , which is small in comparison. We discuss the relative contributions of the terms in Eq. (1b) in more detail below. In what follows, we consider the parameters in Table I. For these, the active gyrotactic mechanism dominates over the passive one.

### C. Full 3D model

In our 3D model, the swimmer can steer around both  $\mathbf{q}$  and  $\mathbf{p}$ , and respond to signals in both directions with the threshold values  $\Delta u^{\text{th}}$  for velocity differences and  $S^{\text{th}}$  for fluid gradients. Our reinforcement learning shows that when only one signal is used, the dominant signal is still  $p_z$  and the strategy (11) remains optimal. There also exists an equivalent strategy where  $q_z$  is the signal and the agent steers around  $\mathbf{p}$  instead of  $\mathbf{q}$ . It gives rise to the same mechanism for aligning  $\mathbf{n}$  against gravity as in Fig. 1, but from rotations around the  $\mathbf{p}$  axis with  $q_z$  as the signal.

For swimmers allowed to use two signals when navigating, the best strategies obtained from reinforcement learning use the combined signal of  $p_z$  and  $q_z$ ; see Table II. They are highlighted in Figs. 2(c) and 2(d). A common trend in the DNS and the statistical model is that the best strategies frequently take actions along the diagonal. It is therefore of interest to compare to the strategy of choosing actions along the diagonal in Figs. 2(c) and 2(d). This diagonal strategy is simply a superposition of the two single-signal strategies described above, which can be expressed mathematically as  $\omega^{(s)} = \omega_p^{(s)}\mathbf{p} + \omega_q^{(s)}\mathbf{q}$ , with

$$\omega_p^{(s)} = -\omega^{(s)} \begin{cases} 1 & \text{if } q_z > p_z^{\text{th}} \\ -1 & \text{if } q_z < -p_z^{\text{th}} \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \omega_q^{(s)} = \omega^{(s)} \begin{cases} 1 & \text{if } p_z > p_z^{\text{th}} \\ -1 & \text{if } p_z < -p_z^{\text{th}} \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

Here,  $\omega_q^{(s)}$  is identical to that in Eq. (11) and  $\omega_p^{(s)}$  takes the same expression with  $p_z$  replaced by  $q_z$  and multiplied by a minus sign as a consequence of the relative handedness between the vectors  $\mathbf{n}$ ,  $\mathbf{p}$ , and  $\mathbf{q}$ . Corresponding to our simulations, we have chosen the same threshold level  $p_z^{\text{th}}$  for both  $p_z$  and  $q_z$ , and that the swimmer turns around the axes  $\mathbf{p}$  and  $\mathbf{q}$  at the same angular rate  $\omega^{(s)}$ . We find that the performance of strategy (14) is slightly better than the optimal policy obtained by reinforcement learning; see Table II. This is a consequence of the reinforcement learning getting stuck in local optima. Although we cannot be certain that strategy (14) is the global optimal strategy for our flows, we expect it to be close to optimal and we analyze its dynamics below.

Strategy (14) gives rise to terms that add to the contributions of the two passive gyrotactic terms in  $\omega_p = \Omega_p - \Delta\Omega_p$  and  $\omega_q = \Omega_q - \Delta\Omega_q$  in Eqs. (6e) and (6f), and therefore strengthen the gyrotactic contribution. Using the simplified dynamics (no flow gradients or passive gyrotaxis) used to derive Eq. (13), the alignment in the present case becomes  $\langle n_z \rangle \sim 1 - (1 + \frac{2}{\pi})[p_z^{\text{th}}]^2$  for small  $p_z^{\text{th}}$  and  $\langle n_z \rangle \sim \frac{16\sqrt{2}}{3\pi}(1 - p_z^{\text{th}})^{3/2}$  for  $p_z^{\text{th}}$  close to unity; see Appendix B for details. In contrast to the case of a single signal, the present strategy allows for perfect alignment in quiescent flows if the threshold level is small enough. The alignment has a plateau for small  $p_z^{\text{th}}$  and decreases sharply to zero as  $p_z^{\text{th}}$  approaches unity, similar to Eq. (13) and shown as the dash-dotted line in Fig. 3(a).

For the case of all four signals,  $p_z$ ,  $q_z$ ,  $S_{np}$ , and  $S_{nq}$ , it is harder to read off a simple strategy from our reinforcement learning results. Since the best strategies found do not surpass the results

of strategy (14) (see Table II), we do not expect that the reinforcement learning has converged to the global optimal strategy. This is no surprise: with nine actions and 81 states considered here, the number of possible strategies is enormous, i.e., of the order  $10^{77}$ . We therefore proceed in a different way to find an efficient and interpretable strategy. In the strategy (14), the signals  $p_z$  and  $q_z$  are used to strengthen the passive gyrotaxis. In Ref. [25],  $S_{np}$  was used to increase the shape factor  $\Lambda$  to a larger effective value, making the swimmers preferentially sample up-welling regions, which facilitates upward migration. This was achieved by choosing  $\omega_q^{(s)} = \omega^{(s)} \text{sign}(S_{np})$  when  $|S_{np}| > S^{\text{th}}$ , and 0 otherwise, i.e., adding a contribution to the angular velocity with the same sign as  $S_{np}$ , effectively increasing  $\Lambda$  by  $\omega^{(s)}$  when  $|S_{np}|$  is above the threshold value. We choose a strategy that is a superposition of the two effects of strengthening the passive gyrotactic angular velocity and increasing the effective shape factor. Testing different combinations of putting priority to either  $\{p_z, q_z\}$ ,  $\{S_{np}, S_{nq}\}$ , or a combination thereof shows that prioritizing  $\{p_z, q_z\}$  works best, i.e., use the strategy (14) when the  $p_z$  or  $q_z$  signal is above its threshold and refine this strategy using  $\{S_{np}, S_{nq}\}$  when the signal is below the threshold,

$$\omega_p^{(s)} = -\omega^{(s)} \begin{cases} 1 & \text{if } q_z > p_z^{\text{th}} \\ -1 & \text{if } q_z < -p_z^{\text{th}} \\ 1 & \text{if } |q_z| < p_z^{\text{th}} \text{ and } S_{nq} > S^{\text{th}} \\ -1 & \text{if } |q_z| < p_z^{\text{th}} \text{ and } S_{nq} < -S^{\text{th}} \\ 0 & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\omega_q^{(s)} = \omega^{(s)} \begin{cases} 1 & \text{if } p_z > p_z^{\text{th}} \\ -1 & \text{if } p_z < -p_z^{\text{th}} \\ 1 & \text{if } |p_z| < p_z^{\text{th}} \text{ and } S_{np} > S^{\text{th}} \\ -1 & \text{if } |p_z| < p_z^{\text{th}} \text{ and } S_{np} < -S^{\text{th}} \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

We did not find any strategy performing better than Eq. (15) using reinforcement learning. We remark that a similar strategy was found in Fig. 3(d) in Ref. [25] for the signals  $p_z$  and  $S_{np}$  in two-dimensional frozen flows. But in that case, the strain signal  $S_{np}$  was prioritized in the optimal strategy, and the  $S_{np}$  signal also dominated over  $p_z$  for the case of a single signal. The explanation for this apparent contradiction is that swimmers that follow the strategy (11) in two-dimensional frozen flows for the parameters considered in Ref. [25] tend to end up at stable fixed points or limit cycles, prohibiting them from upward migration unless a significant amount of Brownian noise is added. This trapping is similar to gyrotactic trapping observed for naive swimmers without steering in shear flows [34,61,62] and vortical flows [63]. In chaotic frozen 3D flows or if the flow fluctuates quickly enough due to turbulence, this trapping goes away and it is then beneficial to adopt the strategies for active gyrotactic stability considered here.

Figure 4 compares the robustness of the found strategies to changes in the flow for a range of  $\tau^{(\eta)}$  with  $u^{\text{rms}} = 1$  mm/s (upper row) and  $u^{\text{rms}} = 10$  mm/s (lower row). The case  $u^{\text{rms}} = 1$  mm/s can, for example, be obtained in a seasonal thermocline, where  $\tau^{(\eta)} \sim 3\text{--}30$  s and the corresponding Taylor scale Reynolds number is relatively small,  $\text{Re} \sim 7\text{--}70$  [42]. The case  $u^{\text{rms}} = 10$  mm/s can be obtained in more turbulent environments, for example, a fjord with  $\tau^{(\eta)} \sim 0.3\text{--}1$  s and  $\text{Re} \sim 70\text{--}200$  [42,64]. The data in Fig. 4 are generated using the statistical model with simulation units such that  $u^{\text{rms}}$  and  $\tau^{(\eta)}$  take their desired values in comparison to the parameters in Table I. Figures 4(a) and 4(d) show the average vertical velocity component  $\langle v_z \rangle$ . Using Eq. (1b), this can be decomposed as

$$\langle v_z \rangle = \langle u_z \rangle + \langle v_z^{(g)} \rangle + v^{(s)} \langle n_z \rangle. \quad (16)$$

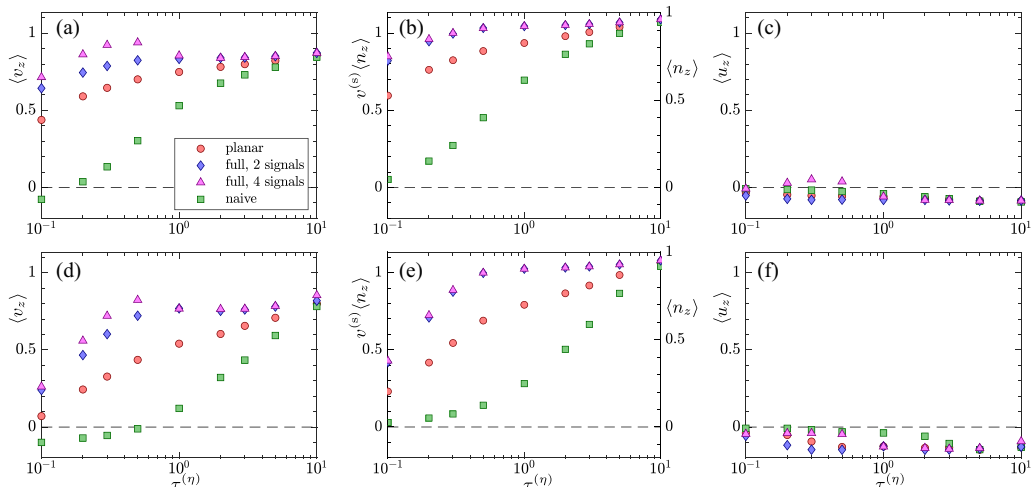


FIG. 4. (a), (d) Steady-state results from statistical model simulations for the average vertical velocity  $\langle v_z \rangle$  against the Kolmogorov time  $\tau^{(n)}$  (units of seconds) for a swimmer following either the planar strategy (11) (red  $\circ$ ), the full 3D strategy using two signals (14) (blue  $\diamond$ ), the full 3D strategy using four signals (15) (magenta  $\triangle$ ), or the naive passively gyrotactic strategy (green  $\square$ ). Parameters of the swimmer are given in Table I. The flow velocity is  $u^{\text{rms}} = 1$  mm/s (top row) and  $u^{\text{rms}} = 10$  mm/s (bottom row) with a constant Kubo number,  $\text{Ku} = 10$ . (b), (e) Same for the vertical velocity due to swimming,  $v^{(s)}\langle n_z \rangle$  (the scale on the right axis shows the alignment  $\langle n_z \rangle$ ). (c), (f) Same for the vertical component of the flow velocity along swimmer trajectories,  $\langle u_z \rangle$ . The displayed velocities are normalized by  $\langle v_z \rangle_0$  in Eq. (10).

Figure 4 shows the two contributions  $v^{(s)}\langle n_z \rangle$  [Figs. 4(b) and 4(e)] and  $\langle u_z \rangle$  [Figs. 4(c) and 4(f)] to  $\langle v_z \rangle$ , normalized by the vertical velocity in a quiescent flow, given by Eq. (10). We do not show the contribution  $\langle v_z^{(g)} \rangle / \langle v_z \rangle_0$  because it is approximately constant and equal to  $-0.1$ .

First, we compare the planar strategy using  $\omega^{(s)} = 0$  (green  $\square$ ) to the naive strategy using  $\omega^{(s)} = 0$  (green  $\square$ ). Figures 4(a) and 4(d) show that the planar strategy has larger average upward velocity  $\langle v_z \rangle$  for all considered flow parameters, only being close for very weak flow gradients,  $\tau^{(n)} \sim 10$  s. While the naive strategy has small or even negative values of  $\langle v_z \rangle$ , the planar strategy is of the order of the vertical velocity in quiescent flows,  $\langle v_z \rangle_0$ , for all considered parameter values, except for the case of strong flow velocity  $u^{\text{rms}} = 10$  mm/s and strong flow gradients,  $\tau^{(n)} \sim 0.1$  s. The explanation for the difference in velocity between the two strategies is the additional alignment  $\langle n_z \rangle$  of the planar strategy due to the active gyrotactic stability; see Figs. 4(b) and 4(e). Even though the planar strategy obtains a slightly more negative sampling of the vertical flow component,  $\langle u_z \rangle$ , than the naive case [see Figs. 4(c) and 4(f)], this difference is negligible compared to the difference in the contribution to the velocity due to alignment,  $v_s\langle n_z \rangle$ .

Next, we consider the full 3D model using two signals (14) (blue  $\diamond$ ). In this model, the alignment reaches a plateau, close to perfect alignment,  $\langle n_z \rangle \approx 1$ , unless the flow gradients are strong. The mean upward velocity shows the same trend, but with a plateau somewhat below  $\langle v_z \rangle_0$  due to the negative contribution from sampling of the vertical flow component,  $\langle u_z \rangle < 0$ . Finally, for the full 3D model using four signals (15) (magenta  $\triangle$ ), the alignment  $\langle n_z \rangle$  is approximately the same as for the case of two signals. But the refined response using the strain components leads to a sampling of the vertical flow component  $\langle u_z \rangle$  that is less negative and even positive for some parameter values. The net result is a larger positive vertical velocity than the strategy with two signals, allowing vertical migration that is nearly as efficient as in the quiescent flow for some parameters.

Simulation results with  $u^{\text{rms}} = 100$  mm/s are identical to the data with  $u^{\text{rms}} = 10$  mm/s within numerical precision (not shown). In this limit, the active mechanism only starts failing when the

flow gradients  $\sim 1/\tau^{(n)}$  are much larger than  $\omega^{(s)}$ . In contrast, the naive strategy fails when the flow gradients are much larger than  $1/(2B)$ , which is one order of magnitude smaller than  $\omega^{(s)}$  for the parameters from Table I. We remark that these parameters are a typical sample. Specific plankton species in nature could have other values of  $\omega^{(s)}$  or  $1/(2B)$ , adjusting the relative importance of the active or passive reorientation mechanism compared to the flow gradients. For values of  $u^{\text{rms}}$  much smaller than 1 mm/s, the average alignment approaches unity for all the cases in the displayed range of  $\tau^{(n)}$  (not shown). In conclusion, the found strategies are robust, showing good performance for a large range of flow parameters.

#### IV. CONCLUSIONS

We used reinforcement learning to find robust and efficient strategies for vertical migration of microswimmers. In the simplest case, the swimmer only has access to a single signal  $\Delta u_p$  or, equivalently,  $p_z$ , which allows for detection of the angle between the orientation of the swimmer (its  $p$  axis) and gravity. By actively rotating in the appropriate direction around its  $q$  axis, the swimmer exhibits an active gyrotactic contribution to the angular velocity which is one order of magnitude larger than the passive gyrotactic contribution for typical copepods. The active contribution results in increased upward alignment, leading to vertical migration of the order of that in quiescent flows unless turbulence is very vigorous. If the swimmer in addition is able to sense  $\Delta u_q$ , it can, up to the resolution set by signal thresholds, measure its orientation relative to gravity, leading to yet stronger upward alignment and vertical migration velocity. By using the strain components  $S_{np}$  and  $S_{nq}$ , the swimmer can further increase its migration velocity by exploiting up-welling regions of the flow [25]. Rotating in the opposite direction allows the swimmer, even though it is bottom heavy, to align with gravity, providing an efficient means for downward swimming.

In the passive gyrotactic mechanism, the shape and mass distribution are important. Using active gyrotactic stability, these parameters are not as important, and even spherical swimmers with homogeneous mass distribution ( $\Lambda = 0$  and  $B = \infty$ ) show the same degree of vertical alignment in turbulent flows. For the parameters in Table I, the passive gyrotactic mechanism only becomes relevant close to quiescent flows, where it allows for better upward alignment than what can be obtained by a single signal  $\Delta u_p$ . The active mechanism is robust and efficient over a large range of the flow parameters  $\tau^{(n)}$  and  $u^{\text{rms}}$ . Simulations of the statistical model and DNS give similar results, both when finding the same types of optimal strategies when training in frozen flows and in the evaluation using time-fluctuating flows.

What bearing do our results have for understanding the behavior of plankton in the turbulent ocean? It is believed that light is important for daily migration of plankton. Strong light determines the swimming direction of many plankton species [2]. We propose that the mechanism discovered here may serve as a means for efficient vertical migration of plankton species at low light intensities, and may also serve as a complementary guide in the presence of light. Further experiments are needed to verify whether plankton in nature have evolved to exploit our proposed mechanism.

#### ACKNOWLEDGMENTS

K.G. acknowledges support from Vetenskapsrådet Grant No. 2018-03974\_VR and by a grant from the Knut and Alice Wallenberg Foundation, Grant No. 2019.0079. J.Q. and L.Z. were supported by the National Natural Science Foundation of China (Grant No. 91752205) and the Institute for Guo Qiang of Tsinghua University (Grant No. 2019GQG1012). L.Z. was supported by a collaboration grant from the joint China-Sweden mobility programme (NSFC-STINT) [Grants No. 11911530141 (NSFC) and No. CH2018-7737 (STINT)]. Statistical model simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC).

TABLE III. Values of the training parameters in our reinforcement learning.

	$\gamma$	$\alpha_0$	$\varepsilon_0$	$E^{(\alpha)}$	$E^{(\varepsilon)}$	$T^{\text{eq}}$	$T^{\text{train}}$	$T^{\text{meas}}$
DNS, 1 signal	0.99	0.01	0.005	500	800	50 s	88 s	0.0088 s
DNS, 2 signals	"	0.02	"	"	"	"	"	"
DNS, 4 signals	"	0.15	0.001	"	"	"	"	"
Statistical model, 1 signal	"	0.01	0.005	"	"	112 s	447 s	N/A
Statistical model, 2 signals	"	0.02	"	"	"	"	"	"
Statistical model, 4 signals	"	0.04	0.02	1000	2000	"	"	"

### APPENDIX A: REINFORCEMENT LEARNING TRAINING PARAMETERS

We train the swimmer using reinforcement learning as described in Sec. II. The training parameters are summarized in Table III for the cases where states consist of a single signal, two signals, or four signals. We train in frozen flows obtained from simulations of Eq. (3) in the DNS and the flow (4) with Gaussian statistics (5) in the statistical model. For both flows, units are scaled such that  $u^{\text{rms}} = 6.7$  mm/s and  $\tau^{(n)} = 1$  s. The total number of episodes in a training session was adjusted to reach convergence to approximately optimal strategies. We used  $\sim 1000$  episodes for the DNS and 3000 episodes for the statistical model.

### APPENDIX B: CALCULATION OF ALIGNMENT IN ABSENCE OF FLOW AND PASSIVE GYROTACTIC TORQUE

To derive Eq. (13), we average the long-time limit of  $n_z$  obtained by following the strategy (11),

$$n_z \rightarrow \begin{cases} \sqrt{1 - [p_z^{\text{th}}]^2 - q_{z,0}^2} & \text{if } |p_{z,0}| > p_z^{\text{th}} \\ n_{z,0} & \text{otherwise,} \end{cases} \quad (\text{B1})$$

over uniformly distributed initial angles. To this end, we parametrize the coordinate system  $\mathbf{n}$ ,  $\mathbf{p}$ , and  $\mathbf{q}$  using three angles ( $0 < \theta < \pi$ ,  $-\pi < \varphi < \pi$ , and  $-\pi < \alpha < \pi$ ) as follows:

$$\mathbf{n} = \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}, \quad \mathbf{p} = \cos \alpha \begin{pmatrix} \sin \varphi \\ -\cos \varphi \\ 0 \end{pmatrix} + \sin \alpha \mathbf{n} \times \begin{pmatrix} \sin \varphi \\ -\cos \varphi \\ 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{q} = \mathbf{n} \times \mathbf{p}. \quad (\text{B2})$$

Here the directions  $\mathbf{p}$  and  $\mathbf{q}$  are orthogonal to  $\mathbf{n}$ , chosen such that  $p_z = 0$  when  $\alpha = 0$  and  $\mathbf{q}$  is chosen to form a right-handed coordinate system. The angles of uniformly distributed directions are distributed according to  $P(\theta, \varphi, \alpha) = \sin \theta / (8\pi^2)$ . Integrating  $\varphi$  away and changing coordinates to  $p_z = -\sin \alpha \sin \theta$  and  $q_z = -\cos \alpha \sin \theta$  gives the joint distribution  $P(p_z, q_z) = 1 / (2\pi \sqrt{1 - p_z^2 - q_z^2})$  of  $p_z$  and  $q_z$  with  $p_z^2 + q_z^2 \leq 1$ . Using this distribution to average the limiting value (B1) for the case  $|p_{z,0}| > p_z^{\text{th}}$  with  $0 \leq p_z^{\text{th}} \leq 1$  over the initial orientation gives (the case  $|p_{z,0}| \leq p_z^{\text{th}}$  gives a zero contribution due to the equal probability of  $n_{z,0}$  taking either sign)

$$\begin{aligned} \langle n_z \rangle &= \int_0^{\sqrt{1 - [p_z^{\text{th}}]^2}} dq_{z,0} \int_{p_z^{\text{th}}}^{\sqrt{1 - q_{z,0}^2}} dp_{z,0} 4P(p_{z,0}, q_{z,0}) \sqrt{1 - [p_z^{\text{th}}]^2 - q_{z,0}^2} \\ &= \frac{2}{\pi} \int_0^{\sqrt{1 - [p_z^{\text{th}}]^2}} dq_{z,0} \sqrt{1 - [p_z^{\text{th}}]^2 - q_{z,0}^2} \operatorname{atan} \left[ \frac{1}{p_z^{\text{th}}} \sqrt{1 - [p_z^{\text{th}}]^2 - q_{z,0}^2} \right]. \end{aligned} \quad (\text{B3})$$

Here we used symmetry of the integrand to only consider positive values of  $p_{z,0}$  and  $q_{z,0}$ , and the limits of integration are obtained from the conditions  $|p_{z,0}| > p_z^{\text{th}}$  and  $p_{z,0}^2 + q_{z,0}^2 \leq 1$ . The remaining integral can be represented using a generalized hypergeometric function, giving Eq. (13).



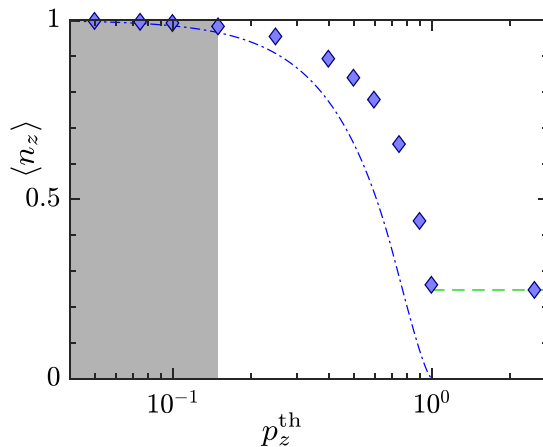


FIG. 5. Same as Fig. 3(a), but for the full 3D model. Symbols show numerical results following the strategy in Eq. (14). The dash-dotted line shows the theory: Eq. (B5) for  $p_z^{\text{th}} < 1/\sqrt{2}$  and two times the value in Eq. (13) if  $1/\sqrt{2} < p_z^{\text{th}} \leq 1$ .

For the case of the strategy using two signals, given by Eq. (14), the long-time limit of  $n_z$  becomes

$$n_z \rightarrow \begin{cases} \sqrt{1 - [p_z^{\text{th}}]^2 - q_{z,0}^2} & \text{if } |p_{z,0}| > p_z^{\text{th}} \text{ and } |q_{z,0}| \leq p_z^{\text{th}} \\ \sqrt{1 - [p_z^{\text{th}}]^2 - p_{z,0}^2} & \text{if } |p_{z,0}| \leq p_z^{\text{th}} \text{ and } |q_{z,0}| > p_z^{\text{th}} \\ \sqrt{1 - 2[p_z^{\text{th}}]^2} & \text{if } |p_{z,0}| > p_z^{\text{th}} \text{ and } |q_{z,0}| > p_z^{\text{th}} \\ n_{z,0} & \text{otherwise.} \end{cases} \quad (\text{B4})$$

Averaging this expression over uniformly distributed angles, as above, gives that  $\langle n_z \rangle$  is two times the value in Eq. (13) if  $1/\sqrt{2} < p_z^{\text{th}} \leq 1$ , and that

$$\begin{aligned} \langle n_z \rangle = & \sqrt{1 - 2[p_z^{\text{th}}]^2} \left\{ 1 - \frac{4p_z^{\text{th}}}{\pi} \text{acos} \left[ \frac{p_z^{\text{th}}}{\sqrt{1 - [p_z^{\text{th}}]^2}} \right] - \frac{1}{\pi} \text{acos} \left[ \frac{1 - 2[p_z^{\text{th}}]^2 - [p_z^{\text{th}}]^4}{(1 - [p_z^{\text{th}}]^2)^2} \right] \right\} \\ & + \frac{4}{\pi} \int_0^{p_z^{\text{th}}} dp_{z,0} \sqrt{1 - p_{z,0}^2 - [p_z^{\text{th}}]^2} \text{acos} \left( \frac{p_z^{\text{th}}}{\sqrt{1 - p_{z,0}^2}} \right), \end{aligned} \quad (\text{B5})$$

if  $p_z^{\text{th}} < 1/\sqrt{2}$ . Equation (B5) is compared to numerical simulations in Fig. 5. The theory agrees qualitatively with the numerical simulations, similar to the theory in the planar model in Fig. 3(a). We did not find a representation of the last integral in Eq. (B5) in terms of standard functions, but it is straightforward to evaluate the term in a series expansion for small  $p_z^{\text{th}}$ , suitable for the interval  $0 < p_z^{\text{th}} < 1/\sqrt{2}$ . The first few terms in the vertical alignment become

$$\langle n_z \rangle \sim 1 - \left( 1 + \frac{2}{\pi} \right) [p_z^{\text{th}}]^2 + \frac{2}{3} [p_z^{\text{th}}]^3 - \frac{1}{2} [p_z^{\text{th}}]^4 + \frac{8}{15} [p_z^{\text{th}}]^5. \quad (\text{B6})$$

- [1] J. G. Park, M. K. Jeong, J. A. Lee, K. Cho, and O. Kwon, Diurnal vertical migration of a harmful dinoflagellate, *cochlodinium polykrikoides* (dinophyceae), during a red tide in coastal waters of Namhae Island, Korea, *Phycologia* **40**, 292 (2001).
- [2] J. Ringelberg, *Diel Vertical Migration of Zooplankton in Lakes and Oceans* (Springer, Netherlands, 2009).
- [3] T. J. Smayda, Adaptations and selection of harmful and other dinoflagellate species in upwelling systems. 2. Motility and migratory behaviour, *Prog. Oceanog.* **85**, 71 (2010).
- [4] F. G. Schmitt, D. Devreker, G. Dur, and S. Souissi, Direct evidence of tidally oriented behavior of the copepod *eurytemora affinis* in the seine estuary, *Ecol. Res.* **26**, 773 (2011).
- [5] T. Katajisto, M. Viitasalo, and M. Koski, Seasonal occurrence and hatching of calanoid eggs in sediments of the northern baltic sea, *Mar. Ecol. Prog. Ser.* **163**, 133 (1998).
- [6] G. C. Hays, C. A. Proctor, A. John, and A. J. Warner, Interspecific differences in the diel vertical migration of marine copepods: The implications of size, color, and morphology, *Limnol. Oceanogr.* **39**, 1621 (1994).
- [7] R. Schuech and S. Menden-Deuer, Going ballistic in the plankton: Anisotropic swimming behavior of marine protists, *Limnol. Oceanogr.* **4**, 1 (2014).
- [8] O. Pundyak, Possible means of overcoming sedimentation by motile sea-picoplankton cells, *Oceanologia* **59**, 108 (2017).
- [9] A. Sengupta, F. Carrara, and R. Stocker, Phytoplankton can actively diversify their migration strategy in response to turbulent cues, *Nature (London)* **543**, 555 (2017).
- [10] R. Hemmersbach and D. P. Häder, Gravitaxial responses of certain ciliates and flagellates, *FASEB J.* **13**, S69 (1999).
- [11] D. P. Häder, M. Lebert, P. Richter, and M. Ntefidou, Gravitaxis and graviperception in flagellates, *Adv. Space Res.* **31**, 2181 (2003).
- [12] R. Hemmersbach-Krause, W. Briegleb, K. Vogel, and D. P. Häder, Swimming velocity of paramecium under the condition of weightlessness, *Acta Protozool.* **32**, 229 (1993).
- [13] J. R. Strickler and A. K. Bal, Setae of the first antennae of the copepod cyclops scutifer (sars): Their structure and importance, *Proc. Natl. Acad. Sci.* **70**, 2656 (1973).
- [14] J. Yen, P. H. Lenz, D. V. Gassie, and D. K. Hartline, Mechanoreception in marine copepods: Electrophysiological studies on the first antennae, *J. Plankton Res.* **14**, 495 (1992).
- [15] J. Titelman and T. Kjørboe, Motility of copepod nauplii and implications for food encounter, *Mar. Ecol. Prog. Ser.* **247**, 123 (2003).
- [16] T. Kjørboe and A. W. Visser, Predator and prey perception in copepods due to hydromechanical signals, *Mar. Ecol. Prog. Ser.* **179**, 81 (1999).
- [17] W. M. Durham, E. Climent, M. Barry, F. De Lillo, G. Boffetta, M. Cencini, and R. Stocker, Turbulence drives microscale patches of motile phytoplankton, *Nat. Commun.* **4**, 2148 (2013).
- [18] F. De Lillo, M. Cencini, W. M. Durham, M. Barry, R. Stocker, E. Climent, and G. Boffetta, Turbulent Fluid Acceleration Generates Clusters of Gyrotactic Microorganisms, *Phys. Rev. Lett.* **112**, 044502 (2014).
- [19] C. Zhan, G. Sardina, E. Lushi, and L. Brandt, Accumulation of motile elongated micro-organisms in turbulence, *J. Fluid Mech.* **739**, 22 (2014).
- [20] K. Gustavsson, F. Berglund, P. R. Jonsson, and B. Mehlig, Preferential Sampling and Small-Scale Clustering of Gyrotactic Microswimmers in Turbulence, *Phys. Rev. Lett.* **116**, 108104 (2016).
- [21] M. Borgnino, G. Boffetta, F. D. Lillo, and M. Cencini, Gyrotactic swimmers in turbulence: shape effects and role of the large-scale flow, *J. Fluid Mech.* **856**, R1 (2018).
- [22] S. Lovecchio, E. Climent, R. Stocker, and W. M. Durham, Chain formation can enhance the vertical migration of phytoplankton through turbulence, *Sci. Adv.* **5**, EAAW7879 (2019).
- [23] J. Qiu, W. Huang, C. Xu, and L. Zhao, Swimming strategy of settling elongated micro-swimmers by reinforcement learning, *Sci. China: Phys., Mech. Astron.* **63**, 284711 (2020).
- [24] J. Qiu, Z. Cui, E. Climent, and L. Zhao, Fluid inertia is an effective gyrotactic mechanism for settling elongated micro-swimmers, [arXiv:2011.10253](https://arxiv.org/abs/2011.10253).
- [25] J. Qiu, N. Mousavi, K. Gustavsson, C. Xu, B. Mehlig, and L. Zhao, Navigation of micro-swimmers in steady flow: The importance of symmetries, *J. Fluid Mech.* (to be published).
- [26] S. I. Dodson, R. Tollrian, and W. Lampert, Daphnia swimming behavior during vertical migration, *J. Plankton Res.* **19**, 969 (1997).

- [27] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, Flow Navigation by Smart Microswimmers via Reinforcement Learning, *Phys. Rev. Lett.* **118**, 158004 (2017).
- [28] K. Gustavsson, L. Biferale, A. Celani, and S. Colabrese, Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning, *Eur. Phys. J. E* **40**, 110 (2017).
- [29] J. K. Alageshan, A. K. Verma, J. Bec, and R. Pandit, Machine learning strategies for path-planning microswimmers in turbulent flows, *Phys. Rev. E* **101**, 043110 (2020).
- [30] L. Biferale, F. Bonaccorso, M. Bucciotti, P. C. D. Leoni, and K. Gustavsson, Zermelo's problem: Optimal point-to-point navigation in 2d turbulent flows using reinforcement learning, *Chaos* **29**, 103138 (2019).
- [31] E. Schneider and H. Stark, Optimal steering of a smart active particle, *Europhys. Lett.* **127**, 64003 (2019).
- [32] S. Muiños-Landin, A. Fischer, V. Holubec, and F. Cichos, Reinforcement learning with artificial microswimmers, *Sci. Robot.* **6**, (2021).
- [33] P. Gunnarson, I. Mandralis, G. Novati, P. Koumoutsakos, and J. O. Dabiri, Learning efficient navigation in vortical flow fields, *Nat. Commun.* **12**, 7143 (2021).
- [34] M. Cencini, G. Boffetta, M. Borgnino, and F. D. Lillo, Gyrotactic phytoplankton in laminar and turbulent flows: A dynamical systems approach, *Eur. Phys. J. E* **42**, 31 (2019).
- [35] G. B. Jeffery, The motion of ellipsoidal particles immersed in a viscous fluid, *Proc. R. Soc. London A* **102**, 161 (1922).
- [36] S. Kim and S. J. Karrila, *Microhydrodynamics: Principles and Selected Applications* (Butterworth-Heinemann, Boston, 1991).
- [37] J. O. Kessler, Hydrodynamic focusing of motile algal cells, *Nature (London)* **313**, 218 (1985).
- [38] L. Machiels, Predictability of Small-Scale Motion in Isotropic Fluid Turbulence, *Phys. Rev. Lett.* **79**, 3411 (1997).
- [39] R. S. Rogallo, *Numerical Experiments in Homogeneous Turbulence* (National Aeronautics and Space Administration, California, USA, 1981), Vol. 81315.
- [40] S. B. Pope, *Turbulent Flows* (Cambridge University Press, Cambridge, 2000).
- [41] K. Gustavsson and B. Mehlig, Statistical models for spatial patterns of heavy particles in turbulence, *Adv. Phys.* **65**, 1 (2016).
- [42] H. Yamazaki and K. Squires, Comparison of oceanic turbulence and copepod swimming, *Mar. Ecol. Prog. Ser.* **144**, 299 (1996).
- [43] H. L. Fuchs and G. P. Gerbi, Seascape-level variation in turbulence- and wave-generated hydrodynamic signals experienced by plankton, *Prog. Oceanog.* **141**, 109 (2016).
- [44] J. Titelman, Swimming and escape behavior of copepod nauplii: Implications for predator-prey interactions among copepods, *Mar. Ecol. Prog. Ser.* **213**, 203 (2001).
- [45] F. Carlotti, D. Bonnet, and C. Halsband-Lenk, Development and growth rates of centropages typicus, *Prog. Oceanog.* **72**, 164 (2007).
- [46] T. Knutsen, W. Melle, and L. Calise, Determining the mass density of marine copepods and their eggs with a critical focus on some of the previously used methods, *J. Plankton Res.* **23**, 859 (2001).
- [47] F. J. Millero, C. Chen, A. Bradshaw, and K. Schleicher, A new high pressure equation of state for seawater, *Deep Sea Res. Part A. Oceanogr. Res. Papers* **27**, 255 (1980).
- [48] D. M. Fields and J. Yen, Implications of the feeding current structure of euchaeta rimana, a carnivorous pelagic copepod, on the spatial orientation of their prey, *J. Plankton Res.* **19**, 79 (1997).
- [49] H. Jiang and G. Paffenhöfer, Relation of behavior of copepod juveniles to potential predation by omnivorous copepods: An empirical-modeling study, *Mar. Ecol. Prog. Ser.* **278**, 225 (2004).
- [50] Z. Kabata and G. C. Hewitt, Locomotory mechanisms in caligidae (crustacea: Copepoda), *J. Fish. Res. Bd. Can.* **28**, 1143 (1971).
- [51] T. Kjørboe, E. Saiz, and A. E. Visser, Hydrodynamic signal perception in the copepod acartia tonsa, *Mar. Ecol. Prog. Ser.* **179**, 97 (1999).
- [52] A. Visser, *Small, Wet & Rational: Individual Based Zooplankton Ecology* (Technical University of Denmark, Copenhagen, 2010).
- [53] J.-S. Hwang and R. Strickler, Can copepods differentiate prey from predator hydromechanically? *Zoolog. Stud.* **40**, 1 (2001).

- [54] J. Yen and J. R. Strickler, Advertisement and concealment in the plankton: What makes a copepod hydrodynamically conspicuous? *Invert. Biol.* **115**, 191 (1996).
- [55] D. Fields, in *The Sensory Horizon of Marine Copepods*, in *Copepods: Diversity, Habitat and Behavior*, edited by L. Seuront (Nova Science, Hauppauge, New York, 2014), pp. 157–179.
- [56] D. M. Fields, Orientation affects the sensitivity of acartia tonsa to fluid mechanical signals, *Mar. Biol.* **157**, 505 (2010).
- [57] E. Buskey, P. Lenz, and D. Hartline, Escape behavior of planktonic copepods in response to hydrodynamic disturbances: High speed video analysis, *Mar. Ecol. Prog. Ser.* **235**, 135 (2002).
- [58] C. J. C. H. Watkins and P. Dayan, Q-learning, *Mach. Learn.* **8**, 279 (1992).
- [59] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)* (MIT Press, Cambridge, 1998).
- [60] B. Mehlig, *Machine Learning with Neural Networks* (Cambridge University Press, Cambridge, 2021).
- [61] W. M. Durham, J. O. Kessler, and R. Stocker, Disruption of vertical motility by shear triggers formation of thin phytoplankton layers, *Science* **323**, 1067 (2009).
- [62] F. Santamaria, F. D. Lillo, M. Cencini, and G. Boffetta, Gyrotactic trapping in laminar and turbulent kolmogorov flow, *Phys. Fluids* **26**, 111901 (2014).
- [63] W. M. Durham, E. Climent, and R. Stocker, Gyrotaxis in a Steady Vortical Flow, *Phys. Rev. Lett.* **106**, 238102 (2011).
- [64] A. E. Gargett, T. R. Osborn, and P. W. Nasmyth, Local isotropy and the decay of turbulence in a stratified fluid, *J. Fluid Mech.* **144**, 231 (1984).